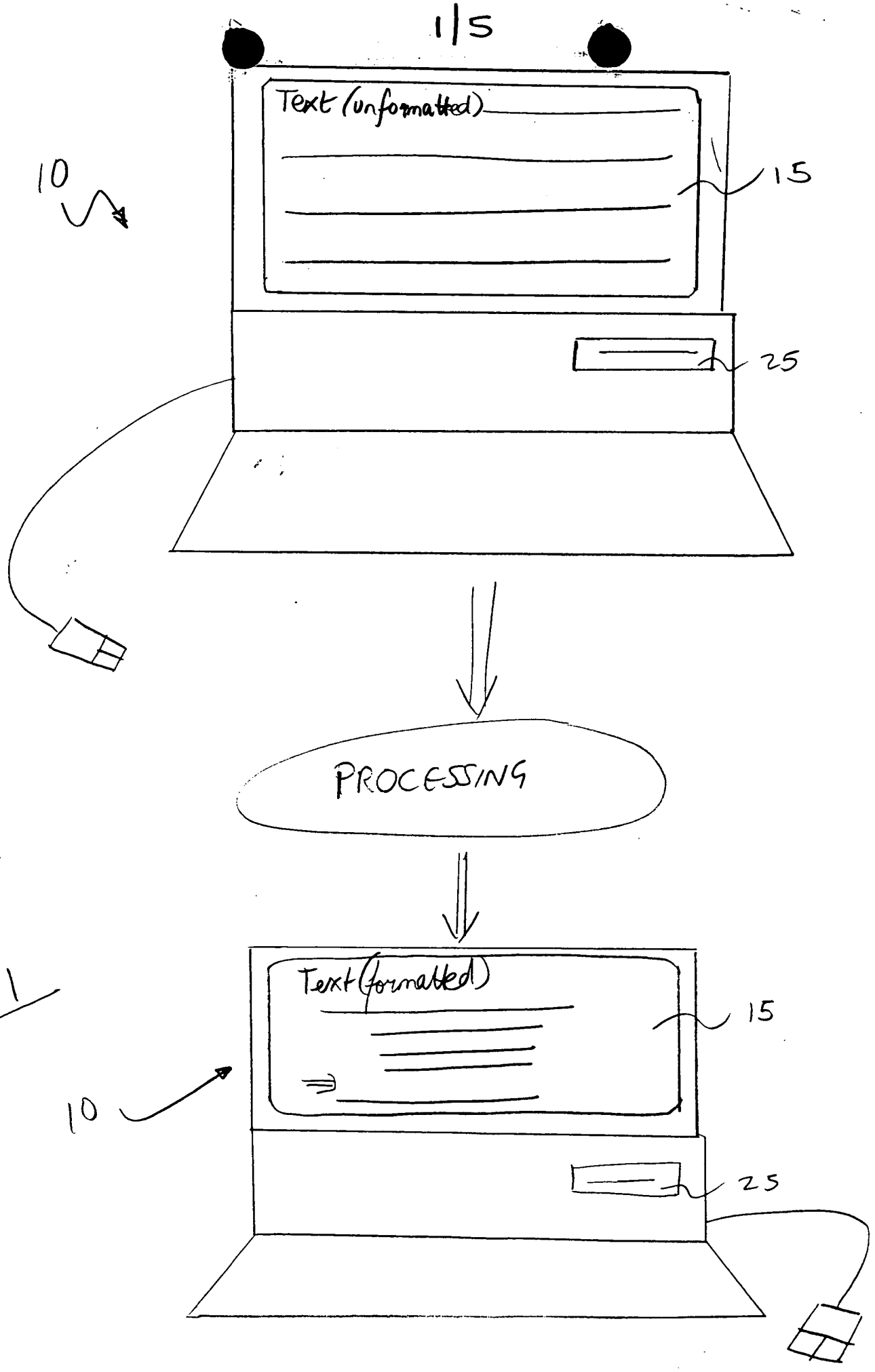
Text (unformatted)
10
15
25
PROCESSING
Text (formatted)
10
15
25

Fig 1

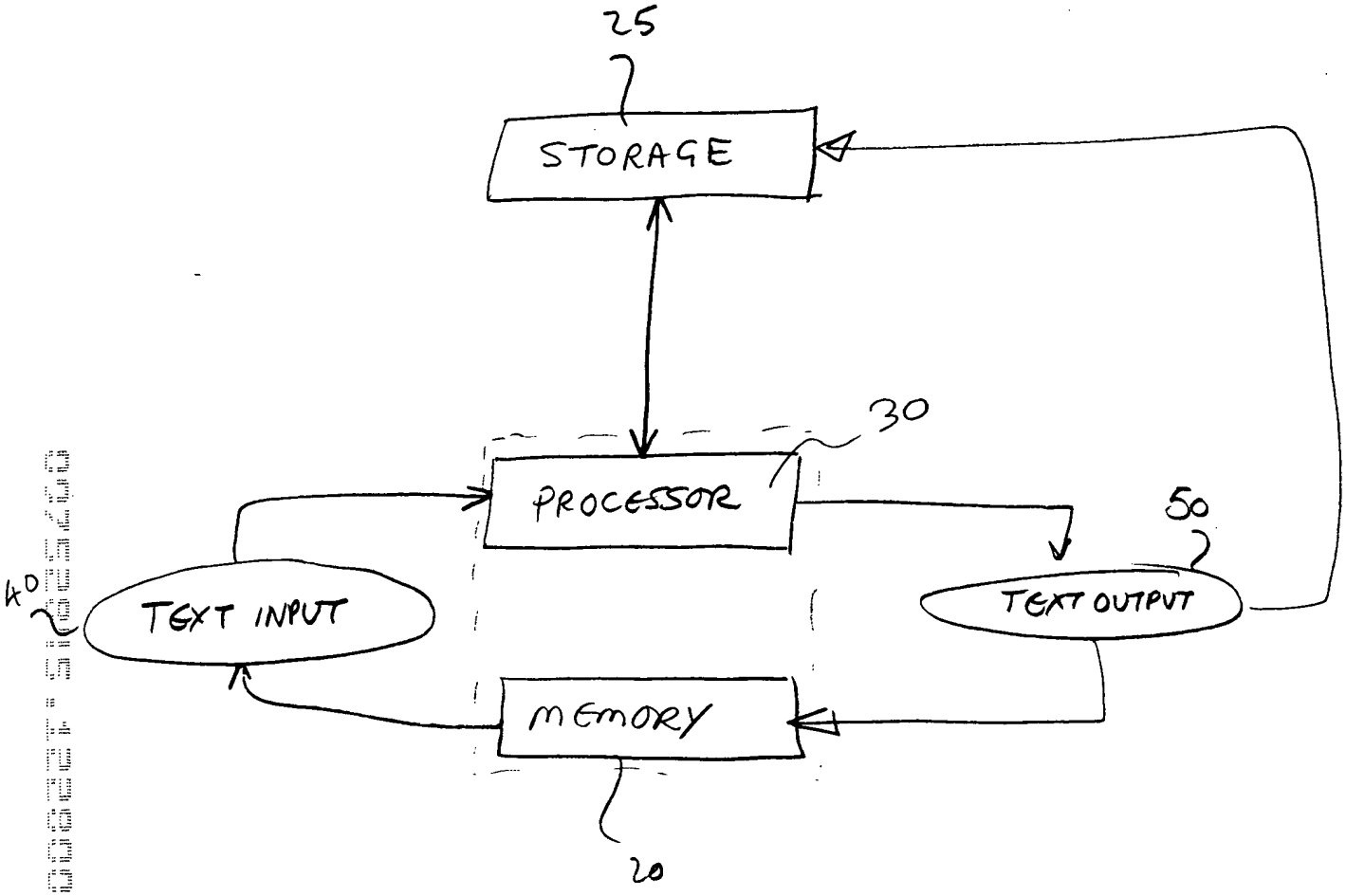
25
STORAGE
30
PROCESSOR
40
TEXT INPUT
50
TEXT OUTPUT
MEMORY
20

FIG. 2

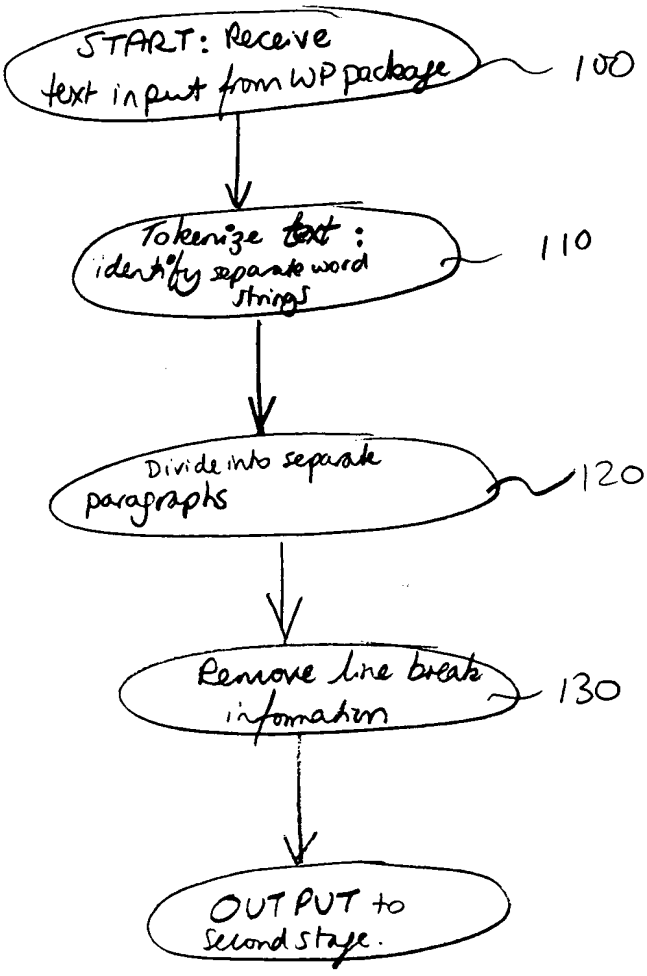
START: Receive text input from WP package
100
Tokenize text: identify separate word strings
110
Divide into separate paragraphs
120
Remove line break information
130
OUTPUT to second stage.

Fig 3

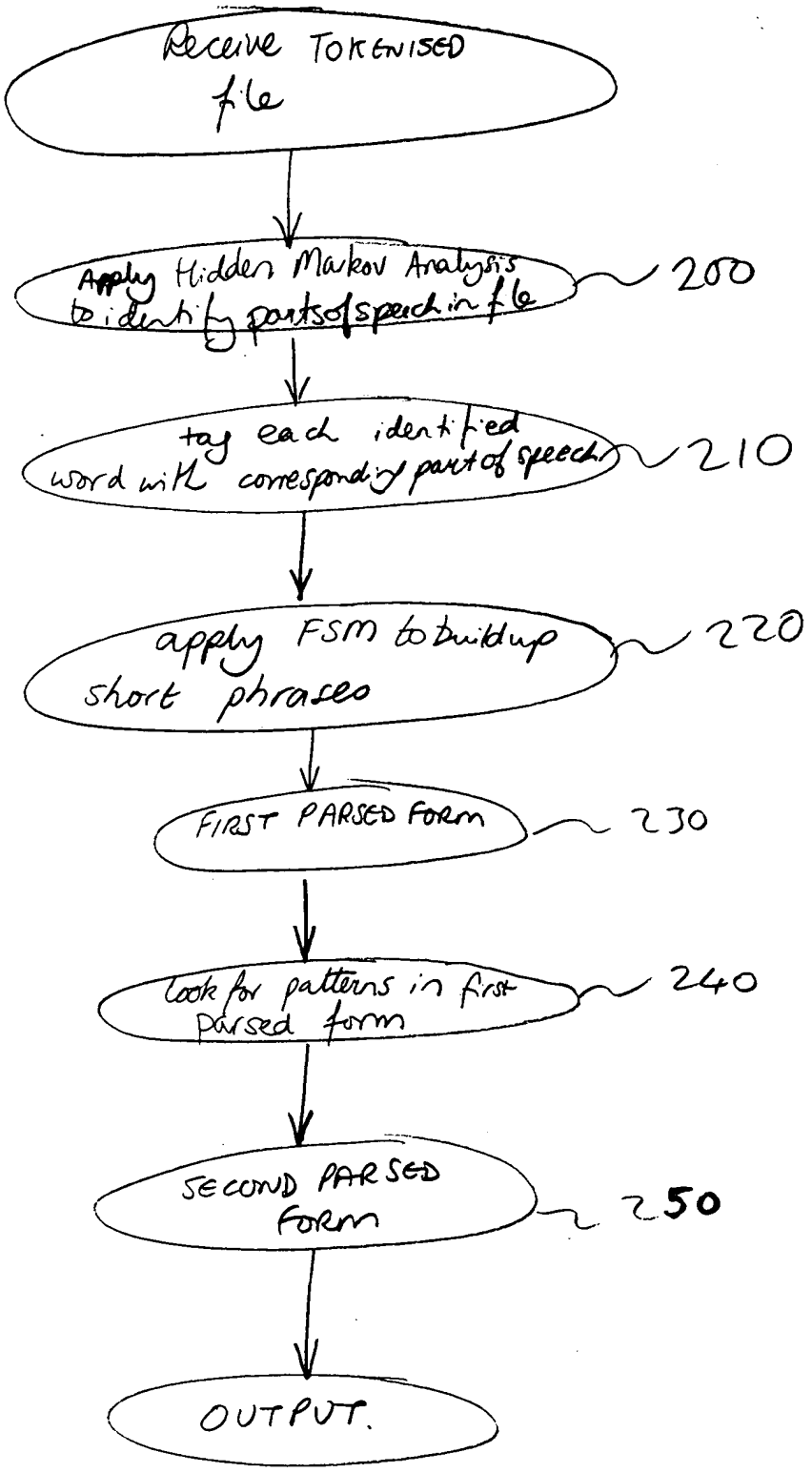
Receive TOKENISED file
Apply Hidden Markov Analysis to identify parts of speech in file
200
tag each identified word with corresponding part of speech
210
apply FSM to build up short phrases
220
FIRST PARSED FORM
230
look for patterns in first parsed form
240
SECOND PARSED FORM
250
OUTPUT.

FIG 4

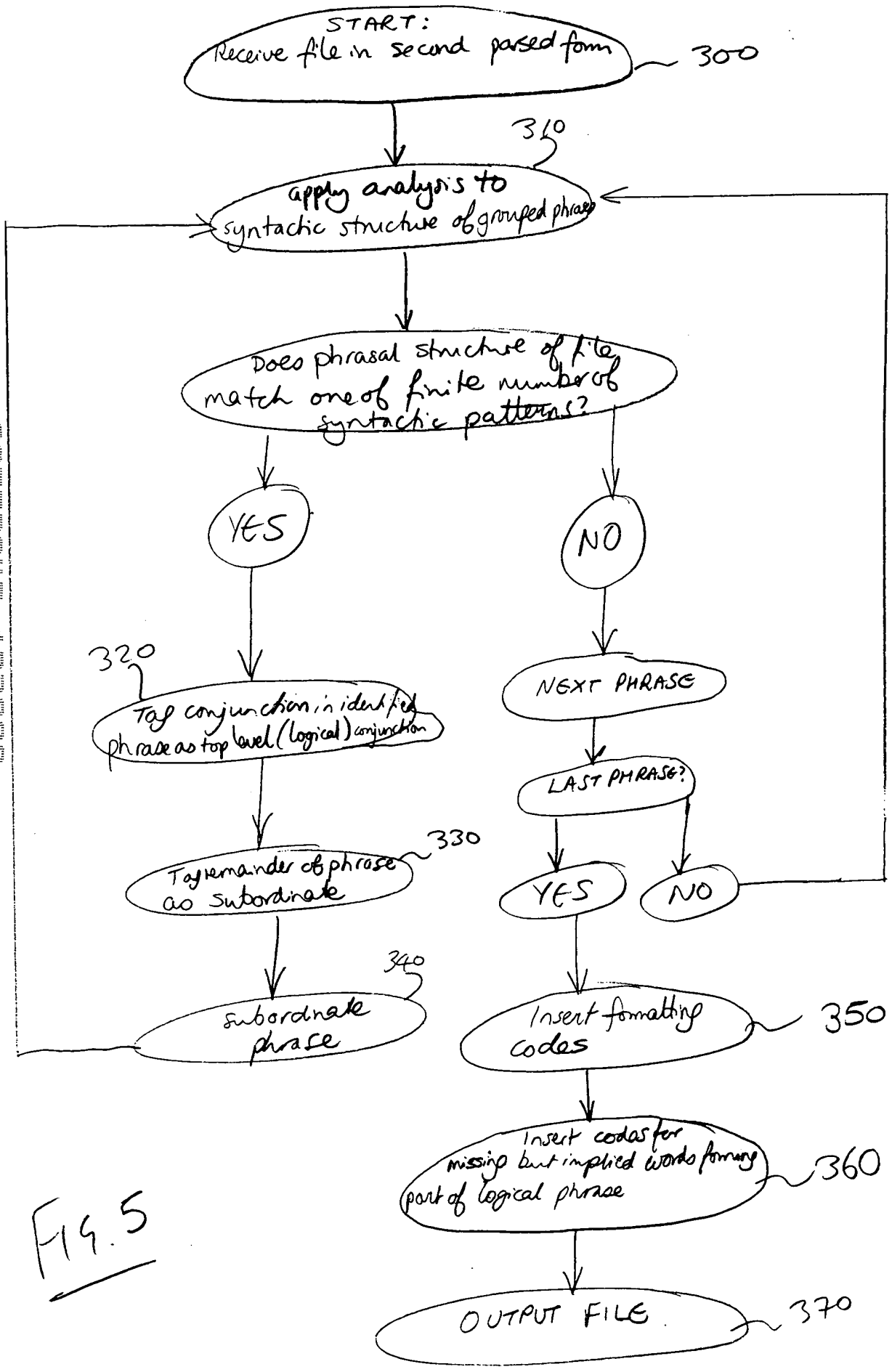
START:
Receive file in second parsed form
300
310
apply analysis to
syntactic structure of grouped phrase
Does phrasal structure of file
match one of finite number of
syntactic patterns?
YES
NO
320
Tag conjunction, if identified
phrase as top level (logical) conjunction
330
Tag remainder of phrase
as subordinate
340
subordinate
phrase
NEXT PHRASE
LAST PHRASE?
YES
NO
Insert formatting
codes
350
Insert codes for
missing but implied words forming
part of logical phrase
360
OUTPUT FILE
370

Fig. 5